



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Risk-adjustment procedures and graphical representations of outcome rates for institutional comparisons

Jacopo Lenzi

Italian Stata Users Group Meeting
November 15, 2018 | Bologna

Introduction

An overriding goal of outcomes research is measuring and comparing hospital performance.

Hospital profiling is a hot topic because provider-specific outcomes, such as mortality rates, are used to:

- Designate centers of excellence
- Determine reimbursement levels in pay for performance programs
- Classify providers as outliers



Risk-adjustment (1)

The outcome rates need to be adjusted to account for pre-existing factors that are outside the quality of care delivered and may confound their assessment.

We all know the direct and indirect methods to calculate age-adjusted rates.

Things get more complicated when other variables, such as clinical factors, are taken into account to derive risk-adjusted rates. These additional characteristics can be retrieved from the hospital discharge abstracts, which are generally inexpensive and enable the analysis of large populations and a large number of clinical conditions.



Risk-adjustment (2)

Statistical techniques to derive hospital-specific risk-adjusted rates can be trivially outlined into three steps:

- Fit an appropriate regression model to the data
- Calculate risk-standardized outcomes resulting from regression analysis
- Display risk-standardized outcomes using appropriate graphical representations

(We are not going to review advanced techniques such as Bayesian hierarchical models.)



Logistic regression

The easiest way to obtain risk-adjusted outcome measures across providers is to build a “conventional” logistic regression model where **depvar** is a binary outcome measure expressed as 0/1 (say death) and covariates **indepvars** are the patient case mix (age, sex, co-morbidities).

```
logit depvar indepvars [if] [in]
```

However, some precautions need to be taken to account for the positive correlation between observations from within the same hospital.



Robust standard errors

When conducting analyses on correlated data, a robust estimation of standard errors should be used. These are known as robust, “sandwich”, or Huber-White standard errors. In this command, ***varname_i*** uniquely identifies providers.

```
logit depvar indepvars [if] [in], vce(cluster varname_i)
```

These estimates allow the correct specification of the mean model while relaxing the assumption of correctly specifying the form of the variance model, i.e., the working correlation matrix.



Generalized estimating equations (GEEs)

Fitting conventional regression models to correlated data might also lead to inefficient parameter estimates.

GEEs, a flexible extension of conventional regression models, account for correlated observations. In Stata, the default GEE within-group correlation structure corresponds to the equal-correlation model, also called “exchangeable”, which is appropriate for profiling studies (Ballinger 2004).

```
xtgee depvar indepvars [if] [in], family(binomial)  
      link(logit) i(varname_i) vce(robust)  
      corr(exchangeable) // c(exc) is optional
```

Robust standard errors should be estimated in conjunction with GEEs (Liang and Zeger 1986) by using the **vce(robust)** option.



Conventional models or GEEs?

Results from GEEs and logistic regression are identical if the within-group correlation is close to zero.

To test whether observations are actually correlated, one should compare a GEE model with an “exchangeable” working matrix and a GEE model with an “independent” working matrix, the latter corresponding to **logit**. The best model has the lowest quasi-likelihood under independence criterion (Pan 2001).

The **qic** command is downloadable from SSC (Cui 2007).

```
qic depvar indepvars [if] [in], family(binomial) link(logit)
    i(varname_i) robust corr(exchangeable)
qic depvar indepvars [if] [in], family(binomial) link(logit)
    i(varname_i) robust corr(independent)
```



Predicted probabilities

After running the best model between the two, we estimate individual risk for each patient (*newvar*) using the observed values of her/his confounding variables.

```
predict newvar [if] [in], rate // , pr in logit post-estimation
```

Then the predicted probabilities can be summed over all records within each provider to derive the expected number of events.

```
tabstat newvar [if] [in], by(varname_i) statistic(sum)
```

The expected number of events is the number of events that would occur if the “standard” event rates had happened, given the provider actual case mix. This can be seen as an “indirect” form of standardization.



Risk-standardized rates

The adjusted outcome for each provider is presented as the ratio of the observed to the expected number of events (O/E ratio). The O/E ratio is favorable if <1 , and unfavorable if >1 .

To allow for comparison of each hospital performance with the national/regional average, the O/E ratio should be multiplied by the overall outcome rate. This measure is called risk-standardized rate. When the events are deaths, the common name is risk-standardized mortality rate (RSMR).

Before discussing how to get these measures in Stata, let us introduce an example based on real data.



Example: 30-day AMI mortality rates in Latvia

We use real data from 20 hospitals in Latvia, year 2016. A total of 2916 patients with acute myocardial infarction (AMI) met the inclusion criteria, and the overall mortality rate was 17.5%.

Death within 30 days of hospital admission for AMI is **Death30Days**, expressed as 0/1, and the hospital identification number is **HospitalID**. Potential confounders include: age, sex, ST elevation status, history of AMI, and 31 co-morbidities based on the Elixhauser method (Quan et al. 2005).

Variables included as confounders in the multivariable model from which RSMRs will be computed are: age, sex, cardiac arrhythmias, pulmonary circulation disorders, peripheral vascular disease, diabetes with complications, and ST elevation status.

... Wait a minute: why were these specific variables selected for inclusion?



Confounder selection

The choice of predictive variables in regression analysis is somewhat of an art. When we use pre-defined sets of covariates, such as Elixhauser's, we may want to avoid model over-fitting and misclassification by including only covariates whose estimated regression coefficients are stable. This can be achieved via numerous automated selection procedures.

See Austin and Tu (2004) for methodological details, and Royston and Sauerbrei (2009) to see how these automated selections can be performed in Stata.



Choosing the right model (1)

We first calculate the QIC value for the exchangeable correlation structure, then we do the same for the independent correlation structure.

```
. set matsize 11000 // default matrix size must be augmented first
. xi: qic Death30Day Sex i.AgeCL4 CARDARRH PULMCIRC PERIVASC DMCX i.AMItype, ///
> family(binomial) link(logit) i(HospitalID) corr(exchangeable) ///
> nolog nodisplay // GEE model with nolog and nodis options to save space
i.AgeCL4          _IAgeCL4_1-4          (naturally coded; _IAgeCL4_1 omitted)
i.AMItype         _IAMITYPE_1-3         (naturally coded; _IAMITYPE_1 omitted)
```

QIC and QIC_u

```
-----
Corr =          exchangeable
Family =        binomial
Link =          logit
p =            11
Trace =         13.999
QIC =           2479.168
QIC_u =         2473.171
-----
```



Choosing the right model (2)

```
. xi: qic Death30Day Sex i.AgeCL4 CARDARRH PULMCIRC PERIVASC DMCX i.AMItype, ///  
> family(binomial) link(logit) i(HospitalID) corr(independent) ///  
> nolog nodisplay // logistic model with nolog and nodis options to save space  
i.AgeCL4          _IAgeCL4_1-4          (naturally coded; _IAgeCL4_1 omitted)  
i.AMItype         _IAMITYPE_1-3         (naturally coded; _IAMITYPE_1 omitted)
```

QIC and QIC_u

```
-----  
Corr =             independent  
Family =           binomial  
Link =              logit  
p =                  11  
Trace =             18.931  
QIC =               2429.175  
QIC_u =             2413.313  
-----
```

The exchangeable structure has a QIC of 2479.168, while the independent structure has a QIC of 2429.175.



Choosing the right model (3)

Conventional logistic regression is the best fitting model here.

```
. xi: logit Death30Day Sex i.AgeCL4 CARDARRH PULMCIRC PERIVASC DMCX i.AMitype, or vce(cluster HospitalID) nolog
i.AgeCL4      _IAgeCL4_1-4      (naturally coded; _IAgeCL4_1 omitted)
i.AMitype     _IAMitype_1-3     (naturally coded; _IAMitype_1 omitted)
```

```
Logistic regression      Number of obs      =      2,916
                        Wald chi2(10)      =      1388.99
                        Prob > chi2       =      0.0000
Log pseudolikelihood = -1195.6566      Pseudo R2         =      0.1155
```

(Std. Err. adjusted for 20 clusters in HospitalID)

Death30Days	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
Sex	1.046859	.1123145	0.43	0.669	.8483312	1.291848
_IAgeCL4_2	2.342332	.3945279	5.05	0.000	1.683749	3.258514
_IAgeCL4_3	3.72511	.8730198	5.61	0.000	2.353156	5.986949
_IAgeCL4_4	7.319825	1.821989	8.00	0.000	4.493935	11.9227
CARDARRH	1.835576	.4022704	2.77	0.006	1.19462	2.820428
PULMCIRC	2.372894	.8879484	2.31	0.021	1.139606	4.940855
PERIVASC	1.802321	.2482871	4.28	0.000	1.375849	2.360985
DMCX	2.315345	.4964455	3.92	0.000	1.520915	3.524734
_IAMitype_2	.4575801	.0678873	-5.27	0.000	.3421225	.6120018
_IAMitype_3	1.312559	.253742	1.41	0.159	.8985984	1.91722
_cons	.0513947	.0109805	-13.89	0.000	.033811	.0781229

Note: _cons estimates baseline odds.

Hospital-specific RSMRs following AMI (1)

After logit post-estimation (**p_hat**), a few command lines have to be run to derive hospital-specific RSMRs and save these values as a new data file.

First, use **collapse** to manipulate data at hospital level and to derive the total number of observed deaths (**Obs**), expected deaths (**Exp**), and patients (**N**)—remember to launch **preserve** first! For each hospital, you generate a new variable (**MR**) containing the crude mortality rates (%).

```
. predict p_hat, pr

. preserve

. collapse (sum) Obs = Death30Day Exp = p_hat (count) N = Death30Day, ///
> by(HospitalID)

. generate MR = Obs*100/N
```



Hospital-specific RSMRs following AMI (2)

After **tabstat**, we define a scalar (**Rate**) containing the overall mortality rate value, that is then used to calculate the RSMRs. We also compute the 95% CIs for RSMRs based on the formula that relates the chi-squared distribution and the Poisson distribution (Garwood 1936).

```
. tabstat Obs N, statistic(sum) save
```

stats	Obs	N
sum	510	2916

```
. matrix total = r(StatTotal)
```

```
. scalar Rate = total[1,1]*100/total[1,2]
```

```
. generate RSMR = Obs/Exp*Rate
```

```
. generate lb_RSMR = (invchi2(2*Obs,0.05/2)/2)/Exp*Rate // RSMR lower limit  
(1 missing value generated)
```

```
. generate ub_RSMR = (invchi2(2*Obs+2,1-0.05/2)/2)/Exp*Rate // RSMR upper limit
```



Hospital-specific RSMRs following AMI (3)

New results are saved as **RSMR.dta** and visualized with **list**.

```
. save RSMR  
. list
```

HospitalID	Obs	Exp	N	MR	RSMR	lb_RSMR	ub_RSMR
1	6	3.260993	23	26.08696	32.17986	11.80944	70.042
2	5	6.318017	38	13.1579	13.84114	4.494181	32.30061
3	14	9.483574	46	30.43478	25.81895	14.11546	43.31982
4	36	35.05648	234	15.38461	17.96043	12.57927	24.86481
5	3	4.12641	21	14.28571	12.71544	2.62223	37.15992
6	23	20.86314	102	22.54902	19.28106	12.22254	28.93106
7	28	17.96316	116	24.13793	27.26201	18.11541	39.4012
8	12	6.220828	35	34.28571	33.73772	17.43277	58.933
9	4	2.108751	7	57.14286	33.1755	9.039207	84.94244
10	6	4.146946	26	23.07692	25.30495	9.286472	55.07823
11	22	24.82706	157	14.01274	15.49816	9.712611	23.46439
12	4	7.993699	52	7.692307	8.75175	2.384557	22.40796
13	14	12.1716	60	23.33333	20.11699	10.99815	33.75289
14	91	124.8035	709	12.83498	12.75255	10.26756	15.65731
15	0	.3216332	1	0	0	.	200.5932
16	31	25.3009	124	25	21.42932	14.56018	30.4172
17	142	152.1697	849	16.72556	16.32086	13.74692	19.23678
18	17	7.870675	36	47.22222	37.77632	22.0061	60.48354
19	36	30.62052	182	19.78022	20.56235	14.40161	28.46695
20	16	14.37248	98	16.32653	19.47022	11.12891	31.61842

```
. restore // restore original data after preserve
```



Graphical representations of RSMRs

RSMR.dta is now used to produce plots of hospital-specific RSMRs.

Plots useful to display outcome rates are (Spiegelhalter 2005):

- Caterpillar plots, sometimes referred to as “forest plots”
- Funnel plots

Both plots, especially funnel plots, operate a fair distinction between small and large providers, and allow to ascertain whether large deviations from the state average are systematic or due to chance.



Caterpillar plot

The caterpillar plot is a sort of league table where providers are ranked according to a performance indicator and, with the aid of CIs, outlying providers are identified. To avoid data misinterpretation, the providers should never be labeled with their rank, and outlying positions have to be strictly determined using CIs.

Plots of estimates and CIs can be obtained in Stata using **ec1plot**, which is downloadable from SSC (Newson 2003).



Caterpillar plot of RSMRs following AMI (1)

Before launching **ec1plot**, a new variable with the ranking of hospitals (**Rank**) must be created. Hospital #15, with only one AMI patient, has been previously discarded.

```
. sort RSMR // Sorting hospitals according to RSMRs

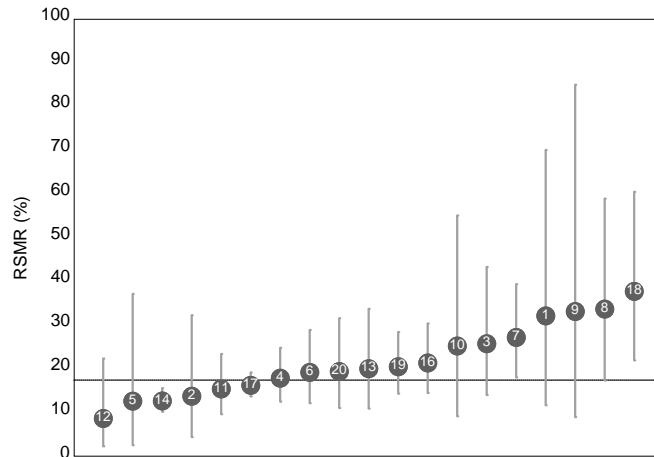
. generate Rank = _n

. ec1plot RSMR lb_RSMR ub_RSMR Rank, plotregion(color(white) ilcolor(black) ///
> margin(none)) graphregion(color(white)) ylabel(0(10)100, angle(360) notick ///
> nogrid) yscale(noextend noline) ytitle("RSMR (%)") xlabel(0 " " 20 " ", ///
> notick) xscale(noextend noline) xtitle("") estopts(mlabel(HospitalID) ///
> mlabposition(0) mlabcolor(white) msymbol(o) msize(huge) mcolor(gs6)) ///
> ciopts(msize(zero) lwidth(medthick) lcolor(gs10)) ///
> baddplot(function y = Rate, col(black) lwidth(thin) range(0 20))
```



Caterpillar plot of RSMRs following AMI (2)

Hospital #14 is a low outlier, while hospitals #7 and #18 are high outliers.



RSMRs following AMI in 19 hospitals in Latvia, Year 2016. Ninety-five per cent confidence intervals are plotted and compared to the overall rate of 17.5%. Hospital #15 is excluded.

Funnel plot

Each hospital-specific RSMR (y axis) is plotted relative to its denominator size (x axis). Given r as the overall rate, n as the hospital volume and z as the standard normal distribution quantile, control limits are superimposed following the formula:

$$y_{\alpha/2}(r, n) = r \pm z_{\alpha/2} \sqrt{\frac{r(1-r)}{n}}$$

where $z_{\alpha/2}$ is 1.96 for 95% control limits and 3.09 for 99.8% control limits.

These boundaries are a measure of precision of the hospital rates, and depend on denominator sizes. Hospitals lying outside the control limits can be seen as outliers.



Funnel plot of RSMRs following AMI (1)

Funnel plots can be obtained either using the **funnelcompar** package (Forni and Gini 2009), or combining a scatter plot with two-way function plots.

We have opted for the second choice. Before doing that, we set highest and lowest values for control limits, based on the range of y axis (up to 60% here).

```
. local up95 = (Rate*(100-Rate))/((60-Rate)/1.96)^2 // Highest 95% limit point
. local low95 = (Rate*(100-Rate))/((0-Rate)/1.96)^2 // Lowest 95% limit point
. local up99 = (Rate*(100-Rate))/((60-Rate)/3.09)^2 // Highest 99.8% limit point
. local low99 = (Rate*(100-Rate))/((0-Rate)/3.09)^2 // Lowest 99.8% limit point
```



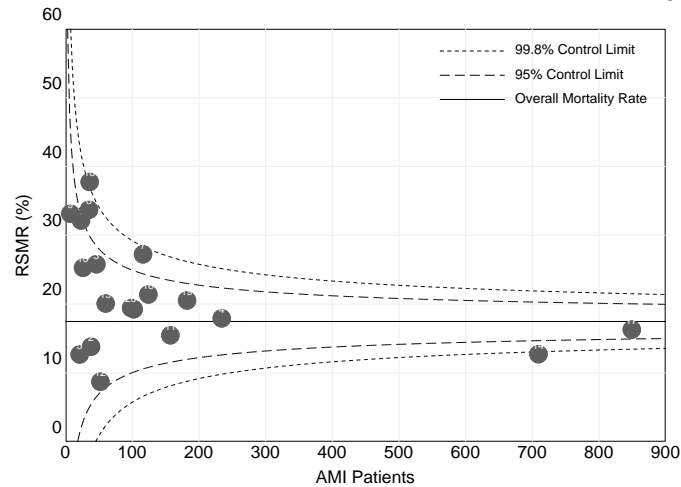
Funnel plot of RSMRs following AMI (2)

The Stata syntax for a customized funnel plot with the range of x axis up to 900 is as follows.

```
. twoway ///
> (function y = Rate+1.96*sqrt((Rate*(100-Rate))/x), col(black) lwidth(thin) ///
> lpattern(dash) range(`up95' 900)) ///
> (function y = Rate-1.96*sqrt((Rate*(100-Rate))/x), col(black) lwidth(thin) ///
> lpattern(dash) range(`low95' 900)) ///
> (function y = Rate+3.09*sqrt((Rate*(100-Rate))/x), col(black) lwidth(thin) ///
> lpattern(shortdash) range(`up99' 900)) ///
> (function y = Rate-3.09*sqrt((Rate*(100-Rate))/x), col(black) lwidth(thin) ///
> lpattern(shortdash) range(`low99' 900)) ///
> (function y = Rate, col(black) lwidth(thin) range(0 900)) ///
> (scatter RSMR N, plotregion(color(white) ilcolor(black) ///
> margin(none)) ytitle("RSMR (%)") xtitle("AMI Patients", height(5)) ///
> ylabel(-0(10)60, angle(360) glcolor(gs15) glwidth(vthin) nogmax nogmin notick) ///
> xlabel(0(100)900, grid glcolor(gs15) glwidth(vthin) nogmax nogmin notick) ///
> xscale(noextend noline) yscale(noextend noline) mlabcolor(black) ///
> mlabsize(vsmall) mfcolor(none) mlcolor(black) mlwidth(thin) bgcolor(white) ///
> graphregion(color(white)) mlabel(HospitalID) mlabposition(0) msymbol(o) ///
> msize(huge) mlcolor(gs8) mfcolor(gs8) mlabcolor(white) ///
> legend(order(3 "99.8% Control Limit" 1 "95% Control Limit" 5 "Overall Mortality Rate") ///
> size(small) col(1) ring(0) pos(2) region(style(none))))
```

Funnel plot of RSMRs following AMI (3)

Hospital #14 is a low outlier, while hospitals #7, #8 and #18 are high outliers.



RSMRs following AMI in 19 hospitals in Latvia, Year 2016. The target is the overall rate of 17.5%. Hospital #15 is excluded.

Caterpillar plots or funnel plots?

Funnel plots might be preferred to caterpillar plots because:

- The eye is naturally drawn to important points that lie outside the funnel
- There is no spurious ranking of institutions
- The relationship of outcome with volume can be preliminarily assessed
- Pairwise comparisons between providers are naturally discouraged



For more details...

... check the next issues of the *Stata Journal*: “Tips for Calculating and Displaying Risk Standardized Hospital Outcomes in Stata” by J. Lenzi and S. Pildava.



References

- Austin, P. C., and J. V. Tu. 2004. Bootstrap Methods for Developing Predictive Models. *The American Statistician* 58: 131–137.
- Ballinger, G. A. 2004. Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods* 7: 127–150.
- Cui, J. 2007. QIC program and model selection if GEE analyses. *Stata Journal* 7: 209–220.
- Forni, S., and R. Gini. Funnel plots for institutional comparisons. *2009 UK Stata Users Group meeting*.
- Garwood, F. 1936. Fiducial Limits for the Poisson Distribution. *Biometrika* 28: 347–442.
- Liang, K.-Y., and S. L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13–22.
- Newson, R. 2003. Confidence intervals and p-values for delivery to the end user. *Stata Journal* 3: 245–269.
- Pan, W. 2001. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics* 57: 120–125.
- Quan, H., V. Sundararajan, P. Halfon, A. Fong, B. Burnard, J. C. Luthi, L. C. Saunders, C. A. Beck, T. E. Feasby, and W. A. Ghali. 2005. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Med Care* 43: 1130–1139.
- Royston, P., and W. Sauerbrei. 2009. Bootstrap assessment of the stability of multivariable models. *Stata Journal* 9: 547–570.
- Spiegelhalter, D. J. 2005. Funnel plots for comparing institutional performance. *Statistics in Medicine* 24: 1185–1202.





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Jacopo Lenzi

Dipartimento di Scienze Biomediche e Neuromotorie
Sezione di Igiene generale ed applicata e Statistica medica

jacopo.lenzi2@unibo.it

www.unibo.it